

Collocation Theory in English And Uzbek From A Contrastive and Corpus-Informed Perspective

 Hamdamov Haydar Azamatovich

PhD in Philology, Acting Associate Professor, Department of Philology, University of Information Technologies and Management, Uzbekistan

Received: 04 February 2026; **Accepted:** 02 March 2026; **Published:** 24 March 2026

Abstract: This article presents a contrastive analysis of the formation and development of collocation theory in English and Uzbek. The study examines collocations from the perspective of habitual co-occurrence in discourse, combinatorial potential, and semantic-functional constraints of lexical units. Major theoretical approaches established in English linguistics—namely the Firthian tradition, systemic functional linguistics, the phraseological school, corpus linguistics, and functional-cognitive frameworks—are critically reviewed in order to identify their explanatory potential and limitations with regard to collocational phenomena in Uzbek. The analysis integrates theoretical discussion, corpus-informed observations, and contrastive methodology to demonstrate that full collocational equivalence between English and Uzbek is relatively rare, while partial equivalence constitutes the dominant pattern. The findings further show that collocational mismatches are not random, but represent systematic phenomena shaped by each language’s internal structure, cognitive modeling, and discourse-specific requirements. On this basis, the study argues for the necessity of contrastive collocational analysis in translation studies and bilingual lexicography. The results of the article provide both theoretical and practical foundations for the comparative study of English and Uzbek collocations, for ensuring collocational adequacy in translation, and for the development of bilingual collocational dictionaries.

Keywords: Collocation, contrastive linguistics, lexical co-occurrence, corpus linguistics, functional-cognitive approach, phraseology, bilingual lexicography.

Introduction: Over the past decades, collocation theory in English linguistics has developed intensively at the intersection of lexicology, phraseology, corpus linguistics, and cognitive linguistics. This has emerged as a key research direction aimed at explaining the language system through so-called “prefabricated units.” Traditional word-centred approaches long dominated linguistic paradigms and treated collocations as a secondary phenomenon. However, contemporary research has demonstrated that actual language use relies to a large extent on stable and semi-stable word combinations. In this respect, Firth’s well-known assertion that “you shall know a word by the company it keeps” provided a methodological

foundation for subsequent studies and marked a decisive shift toward usage-based perspectives (Firth, 1968, p. 179).

Within this line of inquiry, collocations have been conceptualized as units situated between free combinations and idioms—lexico-semantically constrained, yet not fully idiomatic. McKeown and Radev characterize collocations as structures that are partially compositional in meaning but highly restricted in their combinatorial behaviour, drawing particular attention to their automatic identification and practical applicability (McKeown & Radev, 1998, pp. 1-3). This perspective positioned collocations not only as a theoretical construct, but also as a central object of

applied research in fields such as machine translation, text generation, and lexicography.

At the same time, the development of collocation theory in English linguistics has not followed a single linear trajectory. Halliday and Hasan approached collocations as a mechanism of lexical cohesion in discourse, relating them to synonymy and semantic fields within a broader theory of textual organization (Halliday & Hasan, 1976, pp. 274-289). Sinclair drew on extensive corpus evidence and advanced the “idiom principle,” arguing that naturally occurring language is largely organized around prefabricated lexical clusters rather than freely generated combinations (Sinclair, 1991, pp. 110-115). In Russian and Eastern European linguistic traditions, Melchuk’s Lexical Functions model offered a formal and systematic framework for classifying collocational relations within the lexicon (Mel’čuk, 1987, pp. 120-130).

Despite these substantial contributions, much of the existing research tends to explain collocations either primarily through structural-phraseological criteria or through frequency-based statistical measures. As noted by Sonja Poulsen, such approaches often fail to capture the cognitive motivation, prototypical nature, and functional dynamics of collocations in actual language use (Poulsen, 2005, pp. 11-14). In particular, the processes underlying the formation and historical development of collocations within the language system are frequently treated in a fragmentary manner, while their theoretical evolution has rarely been examined within a coherent conceptual framework. Consequently, the logical continuity between the Firthian tradition, structural-phraseological models, corpus-based approaches, and functional-cognitive perspectives remains insufficiently articulated (Poulsen, 2005, pp. 36-38).

Against this background, the present article aims to provide a systematic analysis of the formative stages and developmental trajectories of collocation theory in English linguistics. To achieve this aim, the study pursues the following objectives: (1) to identify early interpretations of the concept of collocation and their methodological foundations; (2) to trace the evolution of collocational analysis across major theoretical schools; (3) to assess the contribution of corpus-based and cognitive approaches to collocation theory; and (4) to evaluate the strengths and limitations of

contemporary integrative models.

The scientific novelty of the study lies in its treatment of collocation theory as a coherent historical-theoretical continuum, highlighting the interconnections between structural, statistical, and cognitive approaches. Collocations are conceptualized not merely as “restricted combinations,” but as cognitive and functional resources of the language system. The theoretical significance of the article consists in synthesizing existing views on collocations within a unified conceptual framework, while its practical significance relates to potential applications in lexicography, translation studies, corpus linguistics, and the teaching of English as a foreign language, particularly with regard to improving methodologies for working with collocational patterns.

LITERATURE REVIEW

The formation and development of collocation theory in English linguistics began with approaches aimed at explaining the habitual co-occurrence of lexical items in discourse and gradually evolved into a theoretical foundation of considerable importance for contrastive and translation-oriented research. In Uzbek linguistics, however, the study of collocations emerged relatively later and largely within the framework of phraseology, while systematic analyses of collocational mismatches between English and Uzbek remain limited. The earliest influential interpretation of collocation in English linguistics is associated with J. R. Firth, who argued that the meaning of a word can be understood through its “habitual companions” (Firth, 1968, p. 179). This perspective emphasizes that combinatorial restrictions of lexical units are determined not primarily by semantic selection, but by patterns of usage and conventionality. This insight is particularly relevant in contrast with Uzbek, where collocational constraints often interact closely with grammatical and morphological factors.

Within the framework of systemic functional linguistics, Halliday and Hasan conceptualize collocations as a mechanism of lexical cohesion contributing to textual coherence (Halliday & Hasan, 1976, pp. 39-42). While this approach has facilitated discourse-level analyses of collocations in English, its applicability to Uzbek remains limited. In Uzbek, textual cohesion is frequently achieved through affixal

mechanisms and syntactic parallelism, which alters the functional load of collocations in comparison with English and reduces the explanatory precision of cohesion-based models. The importance of collocations for lexicographic practice is reflected in the development of specialized collocational dictionaries and reference works (Hausmann, 1989; Benson et al., 1997).

The broader phraseological perspective further situates collocations within a continuum of formulaic language, where fixed expressions, idioms, and semi-fixed combinations interact dynamically in actual language use (Cowie, 1998; Moon, 1998; Wray, 2002). From a usage-based perspective, learner corpus studies have demonstrated that collocations pose persistent challenges even for advanced language users, highlighting the importance of collocational competence beyond grammatical accuracy (Nesselhauf, 2005; Granger & Paquot, 2008). The phraseological approach situates collocations between free combinations and idioms, treating them as units characterized by restricted lexical choice. McKeown and Radev describe collocations as semantically partially compositional yet lexically predetermined structures (McKeown & Radev, 1998, pp. 1-3). From a contrastive perspective, this framework is particularly useful for analyzing structurally divergent but semantically corresponding collocations such as *make a decision* in English and *qaror qabul qilmoq* in Uzbek. Nevertheless, in much of the existing literature such correspondences are often noted at an intuitive level, while systematic modeling of collocational equivalence remains underdeveloped.

A powerful formal apparatus for the classification of collocations is offered by I. Melchuk's Lexical Functions model within the Meaning-Text Theory (Mel'čuk, 1987, pp. 120-130). Although this model has been successfully applied to English, its adaptation to Uzbek has not been sufficiently elaborated. In particular, verb-centred collocations and constructions involving affixal derivation in Uzbek are not fully accommodated within the system of lexical functions, resulting in a methodological gap for contrastive research.

Corpus-based approaches aim to identify collocations through frequency measures and statistical association scores, with Sinclair's "idiom principle" providing a corpus-driven argument for the predominance of

prefabricated lexical units in English (Sinclair, 1991, pp. 110-115). When applied to Uzbek, however, such approaches encounter several limitations. The relatively restricted size and annotation depth of available corpora complicate the reliable identification of collocational patterns. Moreover, the issue that high frequency does not necessarily correspond to genuine collocational status becomes even more acute in contrastive analysis (McKeown & Radev, 1998, pp. 27-30).

Recent corpus-based research published in the *International Journal of Corpus Linguistics* has further refined the understanding of lexical patterning and variability, offering important methodological and empirical insights for collocational analysis. Corpus-based approaches demonstrate that collocations should not be treated as rigid or categorical units, but rather as gradient patterns shaped by frequency, discourse context, and language-specific constraints (Brezina & Meyerhoff, 2024, pp. 345-372). Methodological studies focusing on hypothesis-testing in corpus-assisted discourse research highlight the importance of systematic analytical frameworks for interpreting recurrent co-occurrence patterns beyond surface frequency measures (Marchi, 2025, pp. 401-431). Complementing this perspective, recent corpus-based investigations into emerging combining forms in English illustrate how lexical combinations evolve over time and how corpus data capture both stability and innovation in lexical patterning (Huang & Gao, 2025, pp. 456-498). Taken together, these recent studies reinforce a usage-based and empirically grounded view of collocation, which directly supports the present study's contrastive approach to English-Uzbek collocational relations and its emphasis on partial equivalence and systematic mismatch.

Contemporary functional-cognitive approaches interpret collocations as dynamic and gradient phenomena, linking them to cognitive entrenchment, discourse needs, and semantic framing (Poulsen, 2005, pp. 207-217). This perspective is particularly valuable for explaining collocational mismatches between English and Uzbek, as many lexical restrictions arise not from the linguistic system alone but from broader cultural and conceptual factors. Nevertheless, cognitive explanations are often insufficiently supported by empirical evidence, and experimental or

corpus-based studies addressing English-Uzbek collocational contrasts remain scarce.

A critical review of the existing literature thus reveals several unresolved issues in the contrastive study of collocations in English and Uzbek. First, a shared and operational definition of collocation applicable to both languages has not yet been established (Poulsen, 2005, pp. 221-229). Second, empirical studies offering systematic classifications of collocational equivalence and non-equivalence in translation are limited in number. Third, bilingual lexicographic resources frequently present English-Uzbek collocations in a fragmentary and largely unannotated manner, without explicating their semantic and functional foundations.

For these reasons, future research on English-Uzbek collocations should adopt an integrated methodological framework combining structural-lexical, corpus-statistical, and functional-cognitive approaches. Such an integrative perspective not only advances collocation theory itself, but also offers empirically grounded solutions for translation studies and bilingual lexicography.

RESULTS

The findings indicate that full collocational equivalence between English and Uzbek is relatively rare, while partial equivalence predominates, and structural as well as lexical mismatches display a systematic character. On the basis of corpus-informed and contrastive analysis, the identified collocations were classified into three major groups: (1) collocations exhibiting full semantic and functional equivalence; (2) collocations that are semantically equivalent but lexico-structurally divergent; and (3) cases of collocational mismatch.

Collocations belonging to the first group show a high degree of lexical and semantic correspondence in both languages and therefore require minimal transformation in translation. For instance, the English collocation strong argument and its Uzbek counterpart kuchli dalil share a comparable collocational model in expressing semantic intensity. In such cases, collocational constraints in both languages are grounded in similar cognitive patterns, which allows a high level of equivalence to be preserved in translation.

The second group represents the most frequent type of collocational relation. Here, collocations are semantically corresponding but differ in their lexical composition and structural organization. The English collocation make a decision, for example, is rendered in Uzbek as qaror qabul qilmoq, where the verbal component is not translated directly. The results suggest that collocations formed on the basis of light or auxiliary verbs in English tend to shift toward a noun + verb model in Uzbek. This pattern confirms that collocational constraints are closely tied to language-specific combinatorial properties rather than to semantic correspondence alone (McKeown & Radev, 1998, pp. 1-3).

In the third group, English collocations are not realized as collocational units in Uzbek, but are instead expressed through freer combinations or paraphrastic constructions. For example, while commit a crime functions as a highly stable collocation in English, its Uzbek equivalent jinoyat sodir etmoq, although available, alternates with freer variants in certain contexts. This observation indicates that degrees of collocational stability are not uniform across languages and are shaped by language-specific usage conventions.

Table 1. A contrastive model of English-Uzbek collocations (with frequency information)

No	English collocations	Uzbek equivalent	Degree of equivalence	Frequency in English	Frequency in Uzbek
1	make a decision	<i>qaror qabul qilmoq</i>	Partial equivalence	High	High
2	reach a conclusion	<i>xulosaga kelmoq</i>	Partial equivalence	Medium	Medium

3	pose a question	<i>savol bermoq</i>	Partial equivalence	Medium	High
4	gain experience	<i>tajriba</i> <i>orttirmoq</i>	Partial equivalence	High	Medium
5	take responsibility	<i>mas'uliyatni o'z zimmasiga olmoq</i>	Partial equivalence	Medium	Low
6	provide evidence	<i>dalil keltirmoq</i>	Partial equivalence	High	Medium
7	conduct research	<i>tadqiqot olib bormoq</i>	Partial equivalence	High	High
8	raise an issue	<i>masalani ko'tarmoq</i>	Partial equivalence	Medium	Medium
9	draw attention	<i>e'tiborni qaratmoq</i>	Partial equivalence	High	High
10	achieve results	<i>natijaga erishmoq</i>	Partial equivalence	High	High

Table 1 shows that most English-Uzbek collocational correspondences exhibit partial equivalence, in which

semantic cores are retained while lexico-structural realizations differ systematically across the two languages.

Table 2. Metaphorical and collocational mismatches in English-Uzbek collocations by frequency

No	English collocations	Literal translation	Natural Uzbek equivalent	Frequency in English	Frequency in Uzbek	Note
1	heavy rain	<i>og'ir yomg'ir</i>	<i>kuchli yomg'ir</i>	High	High	Metaphorical shift
2	strong tea	<i>kuchli choy</i>	<i>quyuq choy</i>	Medium	Medium	Different intensity model

3	deep concern	<i>chuqur tashvish</i>	<i>jiddiy tashvish</i>	High	Medium	Spatial metaphor mismatch
4	high probability	<i>baland ehtimol</i>	<i>katta ehtimol</i>	High	High	Quantitative concept
5	quick response	<i>tez javob</i>	<i>zudlik bilan javob</i>	Medium	Medium	Adverbial expansion
6	serious mistake	<i>jiddiy xato</i>	<i>katta xato</i>	High	High	Evaluative adjective difference
7	close relationship	<i>yaqin munosabat</i>	<i>mustahkam munosabat</i>	High	Medium	Physical-abstract shift
8	sharp contrast	<i>o'tkir farq</i>	<i>keskin farq</i>	Medium	High	Metaphorical adaptation
9	weak argument	<i>zaif dalil</i>	<i>asoslanmagan dalil</i>	Medium	Medium	Evaluative strategy
10	broad perspective	<i>keng qarash</i>	<i>keng qamrovli qarash</i>	Medium	Medium	Structural elaboration

Table 2 demonstrates that collocational mismatches frequently occur in high- and medium-frequency adjective-noun combinations, reflecting divergent metaphorical and evaluative preferences in English and Uzbek. The data presented in the tables indicate that differences between English and Uzbek collocations are not random, but systematic and recurrent in nature.

DISCUSSION

Collocational Mapping Model

The findings confirm that theoretical approaches to

collocations developed in English linguistics remain highly relevant for contrastive research. In particular, Firth's notion of collocation as grounded in habitual usage rather than in inherent lexical meaning is also validated in the English-Uzbek comparison: collocational constraints are determined to a greater extent by patterns of conventional use than by the dictionary meanings of individual words (Firth, 1968, p. 179).

To systematically interpret the observed collocational patterns and mismatches between English and Uzbek,

this study proposes a Collocational Mapping Model. The model serves as an integrative analytical framework linking empirical findings with theoretical insights from phraseology, corpus linguistics, and functional-cognitive approaches. Rather than treating collocations as static lexical pairings, the model conceptualizes them as dynamic mappings between semantic cores, language-specific collocational constraints, and target-language realizations. The

proposed model operates in four analytical stages: (1) identification of the semantic core of the source-language collocation; (2) determination of the type of collocational constraint (lexical, metaphorical, or discourse-driven); (3) selection of an appropriate target-language construction; and (4) classification of the resulting relation as full equivalence, partial equivalence, or collocational mismatch.

Figure 1. Collocational Mapping Model

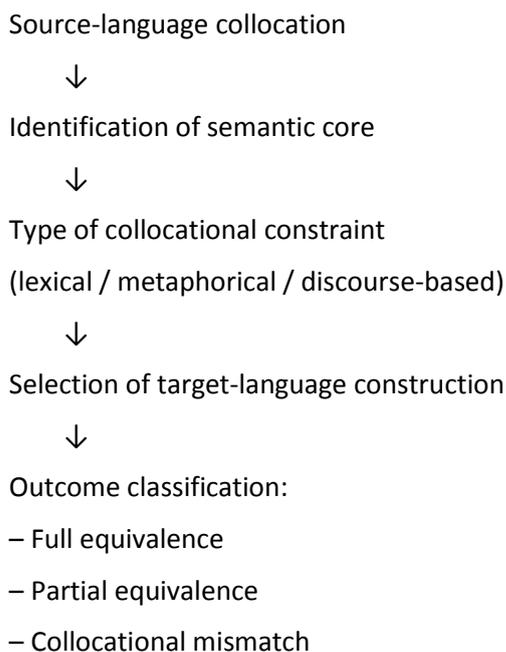


Figure 1. Collocational Mapping Model

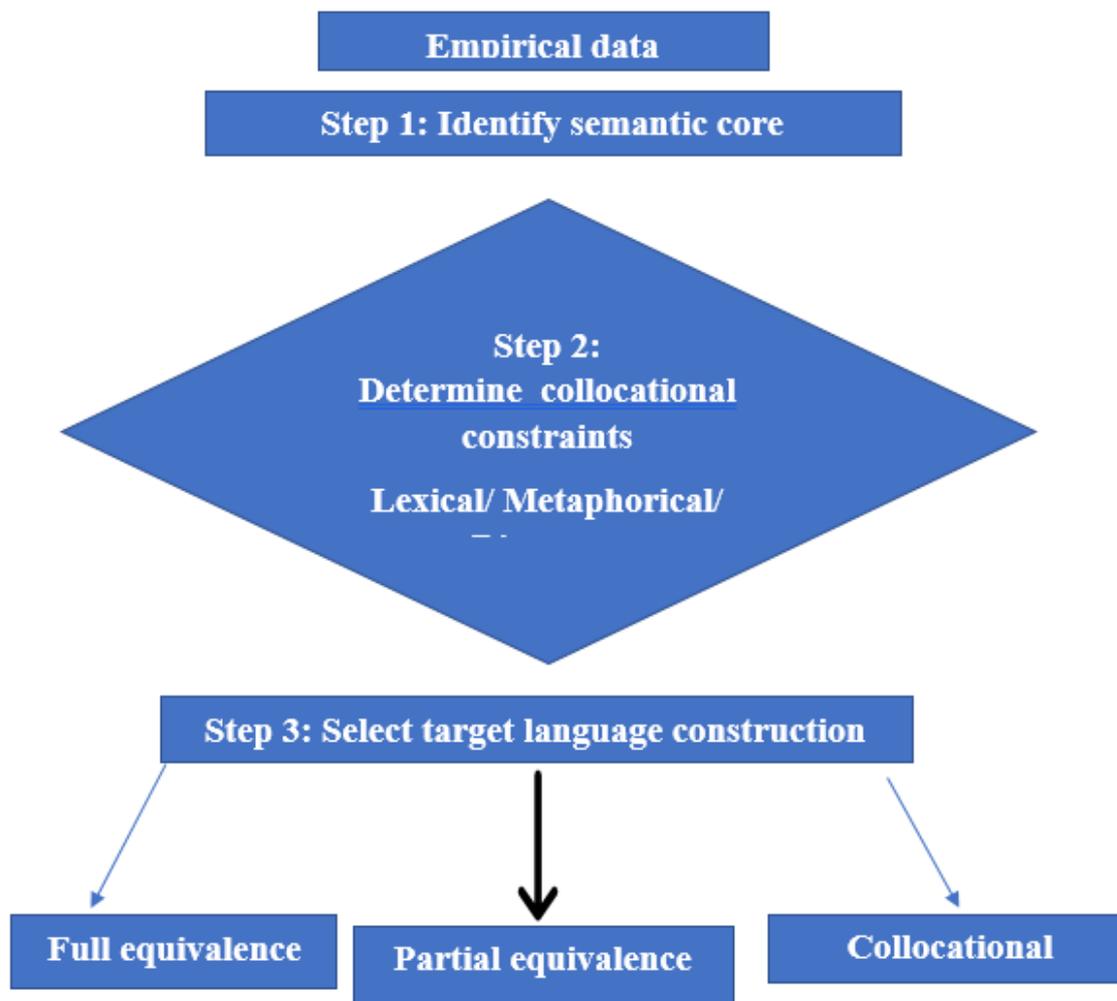


Figure 1 illustrates the proposed Collocational Mapping Model, which outlines the analytical stages used to relate source-language collocations to their target-language realizations in contrastive analysis.

The model is grounded in the empirical patterns presented in Tables 1 and 2 and is not proposed as a universal framework, but as a data-driven analytical tool for contrastive collocational analysis. Conceptually, the model aligns with Firth’s notion of habitual co-occurrence and Sinclair’s idiom principle, while extending these perspectives into a contrastive, translation-oriented framework (Firth, 1968, p.179; Sinclair, 1991, pp. 110-115). It should be noted that the model is currently limited to qualitative and semi-quantitative analysis and does not aim to replace large-scale statistical extraction methods.

Empirical patterns observed in corpus data

Corpus-based inspection of the collocations presented in Tables 1 and 2 reveals a set of stable and recurrent usage patterns that support the contrastive analysis proposed in this study. Across both datasets, English collocations display a high degree of internal stability, particularly in academic and formal discourse, where

verb-noun and adjective-noun combinations such as make a decision, conduct research, provide evidence, heavy rain, and serious mistake recur with consistent semantic and functional profiles. At the same time, the corpus evidence confirms that this stability does not translate into direct lexical or structural equivalence in Uzbek. Instead, Uzbek realizations tend to employ alternative lexico-grammatical strategies, including expanded verbal constructions, paraphrastic reformulations, and shifts in metaphorical motivation.

More specifically, collocations classified as partially equivalent (Table 1) demonstrate that semantic cores are largely preserved across languages, while collocational constraints are reconfigured according to language-specific combinatorial norms. Verb-centred patterns in English frequently correspond to nominally anchored or analytically expanded constructions in Uzbek, indicating a systematic rather than incidental divergence. In contrast, the collocational mismatches

illustrated in Table 2 show that metaphor-driven adjective–noun combinations in English rely on conceptual mappings (e.g. weight, depth, sharpness) that are either absent or non-productive in Uzbek, where alternative evaluative or quantitative strategies are preferred.

Importantly, these mismatches are attested not only in low-frequency or marginal expressions, but also in medium- and high-frequency items, suggesting that collocational non-equivalence constitutes a regular feature of English-Uzbek lexical interaction rather than an exception. This observation supports earlier claims that high frequency alone does not guarantee collocational status and must be interpreted in relation to language-specific constraints (McKeown & Radev, 1998, pp. 27-30).

Theoretical and translational implications

The discussion results indicate that collocational behaviour reflects deeper typological and cognitive differences between English and Uzbek. While Sinclair’s idiom principle operates with high regularity in English, its application in Uzbek appears more flexible, a difference that can be attributed to the agglutinative structure of Uzbek and the role of affixation in encoding semantic relations (Sinclair, 1991, pp. 110-115).

From a functional-cognitive perspective, collocational mismatches should not be interpreted as deficiencies in the lexical system, but rather as manifestations of language-specific conceptualization and discourse needs. For instance, the English collocation heavy rain encodes intensity through a weight-based metaphor, whereas Uzbek prefers a force-based metaphor (*kuchli yomg’ir*), illustrating divergent cognitive framing strategies (Poulsen, 2005, pp. 207-217).

Overall, the findings confirm that a purely structural or statistical approach is insufficient for contrastive collocation studies. Instead, an integrated methodology combining corpus evidence, functional-cognitive interpretation, and comparative semantic analysis is required to adequately account for collocational behaviour in English and Uzbek. Such an approach has direct implications for translation practice and bilingual lexicography, where collocational equivalence must be treated as a matter of semantic and functional adaptation rather than direct lexical transfer. In applied contexts, collocations have been

identified as a key component of lexical competence and an essential focus in language teaching methodologies (Lewis, 2000). These observations are consistent with the lexical priming hypothesis, which argues that repeated exposure to specific co-occurrence patterns shapes speakers’ expectations and constrains lexical choice in discourse (Hoey, 2005). Corpus-based studies of lexical semantics further confirm that meaning emerges from recurrent patterns of co-selection rather than from isolated lexical items, reinforcing the central role of collocations in discourse organization (Stubbs, 2001). From a pedagogical perspective, the explicit teaching of collocations has been shown to enhance fluency and idiomaticity, particularly when supported by corpus-based evidence and pattern-oriented instruction (Lewis, 2000; Hunston & Francis, 2000).

CONCLUSION

This study set out to identify similarities and differences between collocational systems in English and Uzbek through a contrastive analysis of the major conceptual approaches to collocation theory. The findings confirm that collocations constitute a central linguistic phenomenon in both languages, shaped by habitual usage and the combinatorial potential of lexical units. At the same time, the results demonstrate that their structural, semantic, and functional properties do not manifest uniformly across languages. While collocations in English tend to be formed predominantly through light verbs and stable lexical pairings, collocational constraints in Uzbek are more frequently realized through verb-centred constructions, affixal mechanisms, and processes of semantic expansion. This divergence explains why partial equivalence rather than full equivalence characterizes the majority of English-Uzbek collocational correspondences. Moreover, the study establishes that collocational mismatches are not accidental, but systematic phenomena linked to each language’s internal structure, cognitive models, and discourse-specific needs.

The findings corroborate Firth’s claim that collocation is grounded in habitual usage when viewed from a contrastive perspective, while also demonstrating that Sinclair’s “idiom principle” operates with a high degree of regularity in English but manifests in a more flexible form in Uzbek. From a functional-cognitive standpoint,

collocational constraints emerge not merely as lexical or structural phenomena, but as processes closely connected to conceptualization and cultural factors. This perspective proved particularly effective in providing a deeper and more coherent interpretation of collocational behaviour in contrastive analysis.

The scientific significance of the study lies in its proposal of an integrative approach to the contrastive investigation of collocations in English and Uzbek. By combining structural-lexical analysis, corpus-based evidence, and functional–cognitive interpretation, this approach enables a systematic description of collocational equivalence and non-equivalence across languages. As a result, the necessity of studying collocations not only within individual languages but also from a cross-linguistic perspective is theoretically substantiated.

From a practical standpoint, the findings are directly relevant to translation studies and bilingual lexicography. In particular, the results may be applied to the identification of potential collocational mismatches in English-Uzbek translation and to the development of strategies for their adequate rendering. Furthermore, the study provides a methodological foundation for improving the teaching of collocations in English as a foreign language and for offering a more systematic description of collocational units in Uzbek.

Several directions for future research emerge from this study. First, further empirical investigation of English and Uzbek collocations based on large-scale and balanced parallel corpora is required. Second, experimental methods-such as psycholinguistic testing and acceptability judgments-may be employed to examine collocational equivalence and non-equivalence more rigorously. Third, the development of a bilingual lexicographic model for Uzbek-English collocations and its implementation in practical translation contexts represents a promising avenue for future work.

In conclusion, the contrastive study of collocations in English and Uzbek yields significant insights not only for theoretical linguistics, but also for translation practice, lexicography, and language teaching. The framework proposed in this article provides a conceptual and methodological foundation for future research in this

area and may inform future corpus design and bilingual lexicographic practice for under-researched language pairs.

REFERENCES

1. Firth, J. R. (1968). A synopsis of linguistic theory 1930-1955. In F. R. Palmer (Ed.), *Selected papers of J. R. Firth* (pp. 168-205). Longman.
2. Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Longman.
3. Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
4. McKeown, K., & Radev, D. (1998). Collocations. In R. Mitkov (Ed.), *The Oxford handbook of computational linguistics* (pp. 1-29). Oxford University Press.
5. Mel'čuk, I. (1987). *Dependency syntax: Theory and practice*. State University of New York Press.
6. Mel'čuk, I., Chertov, A., & Polguère, A. (1995). Lexical functions as tools for lexicographic description. In L. Wanner (Ed.), *Lexical functions in lexicography and NLP* (pp. 1-20). John Benjamins.
7. Poulsen, S. (2005). *Collocations as discourse phenomena*. Peter Lang.
8. Cowie, A. P. (1998). *Phraseology: Theory, analysis, and applications*. Oxford University Press.
9. Hausmann, F. J. (1989). Le dictionnaire des collocations. In *Wörterbücher / Dictionaries / Dictionnaires* (pp. 1010-1019). De Gruyter.
10. Benson, M., Benson, E., & Ilson, R. (1997). *The BBI dictionary of English word combinations*. John Benjamins.
11. Brezina, V., & Meyerhoff, M. (2024). Corpus-based approaches to lexical patterning and variability. *International Journal of Corpus Linguistics*, 29(3), pp. 345-372.
12. Nesselhauf, N. (2005). *Collocations in a learner corpus*. John Benjamins.
13. Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Blackwell.
14. Granger, S., & Paquot, M. (2008). Disentangling the phraseological web. In *Phraseology: An interdisciplinary perspective* (pp. 27-49). John Benjamins.

15. Moon, R. (1998). Fixed expressions and idioms in English. Oxford University Press.
16. Wray, A. (2002). Formulaic language and the lexicon. Cambridge University Press.
17. Hoey, M. (2005). Lexical priming: A new theory of words and language. Routledge.
18. Huang, J., & Gao, Y. (2025). A corpus-based study into new combining forms in American English. *International Journal of Corpus Linguistics*, 30(4), 456–498. <https://doi.org/10.1075/ijcl.23039.hua>
19. Lewis, M. (2000). Teaching collocation: Further developments in the lexical approach. Language Teaching Publications.
20. Marchi, A. (2025). Hypothesis-testing in corpus-assisted discourse studies: A methodological exploration. *International Journal of Corpus Linguistics*, 30(4), 401–431. <https://doi.org/10.1075/ijcl.24049.mar>
21. Hunston, S., & Francis, G. (2000). Pattern grammar: A corpus-driven approach. John Benjamins.